総合倫理学(synthetic ethics)に向けて

AUTHOR(S):

久木田, 水生

CITATION:

久木田, 水生. 総合倫理学(synthetic ethics)に向けて. 京都大学文学部哲学研究室紀要 2010, 12: 96-105

ISSUE DATE:

2010-02

URL:

http://hdl.handle.net/2433/97994

総合倫理学（synthetic ethics）に向けて

久木田水生

1 序

1.1 目的

本論の目的は、近年 Floridi によって提唱されている「情報倫理」を、メタ倫理学への一つのアプローチとしての「総合的方法（synthetic method）」（以下、SE）に応用することの有用性を検討することである。

総合的方法は心理学や動物行動学の分野で発達してきた方法であり、現実の動物の振る舞いを分析することによってではなく、人工的に作られた環境における人工的行為者の振る舞い、あるいは行為者を含むシステムの一部や全体の振る舞いを観察することによって、その振る舞いについての機械論的な理解を得ることを目指す。

メタ倫理学へのアプローチとしての総合的方法は、行為者の倫理的・道徳的な振る舞いを、その行為者の条件と環境的条件から、機械論的に説明することを目的とする。より具体的に言えばこの方法は、社会がある道徳的行為を促進するにはどのような環境的要因が重要であるか、あるいは特定の要因が存在する環境でどのような道徳的（反道徳的）行動が誘引されやすいか、などの問題に焦点を当てる。

この方法論の基本的なアイディアは、「倫理的」に振る舞う行為者を人工的に作ることによって、道徳性にまつわる様々な問題にアプローチできるだろう、というものである。もちろんここで言われる「倫理的行為」は現実の世界での倫理的行為とは様々な点で異なっている。その行為には自由意思による決定、意図や目的、結果の予測などの要因が欠けているかもしれない。それにもかかわらず、本論で私たちはこのようなモデルを作ることがメタ倫理学の有用な方法でありうることを確認したい。

1.2 背景

倫理学に対する総合的アプローチが提案される背景には、近年の二つの研究の潮流が関係している。一つは MacLennan が「総合動物行動学」と呼ぶ研究分野の発展であり、もう一つは環境倫理などに代表される、行為者よりも行為の影響を受ける対象を関心の中心にする「非標準的」倫理学である(1)。特にここでは、Floridi がコンピュータ倫理の基礎理論として提唱している「情報倫理」に注目する。

総合的方法が意味を持つためには、第一にモデル化したい現象とモデルの間に何ら

かのアナロジーが成り立っていなければならない。第二に、モデルはモデル化したい現象の（当面の目的にとって）重要でない側面をそぎ落としたものでなければならない。したがってモデルを構築する際には、モデル化したい現象の持つ多種多様な側面のどれに焦点を当てるか、ということを考えなければならない。このことは、モデルを構築する際には、私たちは前もって何らかの理論を必要とする、ということを意味する。

ここで問題となるのは、ある出来事（行為）に対して道徳性を帰属させる基準をどのように定めるか、ということである。個人の動機や意図を道徳性の基礎に置く理論は、総合的方法にとって利用可能ではない。というのも、現在のモデルに含まれている人工的行為者に動機や意図を帰属させることが可能か、ということがまず大きな問題になるからである。そこで功利主義や帰結主義が総合的方法の基礎理論として考えられるかもしれない。しかしこれらの理論も十分ではない。総合的方法にとっては可視的な値として道徳性が現れてこなければならないが、これらの理論は道徳性を定量化する方法を提供しないからである。そこで私たちは Floridi の情報倫理を、総合的方法の基礎理論として採用することを検討し、その利点をみていこうと思う。

1.3 論文の構成

以下、2 節では総合動物行動学の方法論とその成果を紹介する。3 節では総合的アプローチを倫理学に応用するための基礎理論として何が必要であるかを考察し、そのような理論の一つとして、Floridi の情報倫理の有用性を検討する。

2 総合動物行動学

人工知能（artificial intelligence）と人工生命（artificial life）はそれぞれ、知性や生命に関する哲学的問題に対して，興味深いアプローチを提供してきた。そのアプローチの特徴は、ある対象／現象を分析する（analyze）ことによって理解しようとするのではなく、それらに相当する対象／現象を人工的に作る（synthesize）ことによって理解しようとすることである。人工知能が本当の意味での知能と呼ぶに値するかどうかに関しては、この分野の誕生から現在に至るまで、活発な議論がなされているが、いまのところ本当の意味で知能と呼べるようなものは生み出されていない。しかしそれにも関わらず、これらの分野は哲学的議論に重要な貢献をしてきた。実際、人工知能は 20 世紀以降、最も豊かな哲学的議論の源を提供した主題の一つである。そして人工知能は、それに関する哲学的議論と共に、「知性とは何か？」という問題に対する重要な示唆を与えてきた(2)。

人工知能や人工生命における試みは、現在ではMacLennanが「総合動物行動学（synthetic ethology）」と呼ぶ、新しい分野を生み出してきた(3)。この分野に従事する研究者たちはコンピュータやロボットを用いたシミュレーションにより、人工的な環境下に置かれた生物や生物集団に特定の振る舞いを創発させることを目指す。従来の人工知能との違いは、あらかじめ一定の振る舞いをするようなプログラムを与えておくのではなく、進化に類比的なプロセスをコンピュータによってシミュレートし(4)、当の振る舞いが環境との相互作用によって生じるメカニズムあるいは原理を明らかにしようとする点にある。

総合動物行動学が対象とする行動・振る舞いには、コミュニケーションや協力行動などが含まれる。総合動物行動学はこれらの動物行動の機械論的な理解に一定の貢献をしてきた。それは、従来の実験室での実験やフィールドワーク、まして思弁的な議論では決して得られなかったであろうような理解である。

本節では総合動物行動学の方法論を検討し、その例を紹介する。

2.1 総合動物行動学の方法論

伝統的に、動物の行動を研究する方法には大きく分けて二つのものがある。一つは実験室のコントロールされた環境での行動を観察する方法（実験的方法）であり、もう一つは実際に生物が生息している環境での行動を観察する方法（フィールドワーク）である。これらの方法はどちらも「分析的方法（analytic method）」と呼ぶことが出来るだろう。というのも、どちらの方法も研究の対象を観察し、その対象の行動を分析的に理解することを試みる方法だからである。

実験的方法は、関連する要因のコントロールと行動の観察が容易であるというメリットを持つ一方で、被験動物が本来の生息環境から切り離されているために、その行動が現実の行動を忠実に反映したものではないかもしれない、というデメリットを持つ。一方フィールドワークは、生物の自然な振る舞いを観察できるというメリットを持つ反面、現実の環境の複雑さが、その行動にとっての本質的な要因を理解しにくくするというデメリットがある。というのも複雑な要因が絡み合う自然環境の中では、ある行動を引き起こした要因を正確に特定することが困難であるし、またある要因だけを変化させて行動の変化を調べることがほとんど不可能だからである。

さらに、より重要なことであるが、両者に共通のデメリットとして、分析的方法では決して理解できない現象がある、ということが挙げられる。たとえば生物進化や言語の発達のような現象は、実験的方法によって再現できるものではないし、フィールドワークによって（現実的な研究のスパンで）観察されるものでもない。このゆえに、

進化論や言語起源論は科学の名に値しない、とさえ言われてきたのである。

総合的方法は、これらの問題に対する解決を提供する。第一に、被験動物は、その「自然な生息地」──それらがそこで育ち、そこに適応して進化してきた環境──での行動を観察される。第二に、環境は完全にコントロールされており、かつ環境の状態や被験動物の内部状態を表すパラメータの値は観察者に対して可視的である。第三に、被験動物のライフサイクルや世代交代もまた私たちのコントロールのもとに置かれている。したがって自然環境では観察することが出来ない進化的なプロセスが観察可能であるし、環境や被験動物の初期設定を様々に変化させて、どのような進化が起こるかを実験的に確かめることが出来る。さらに同じ条件で実験を繰り返すことで、進化の結果のどれが偶然的なもので、どれが本質的なものかを区別することが出来る。

次節では総合的動物行動学の初期のアプローチの一例を紹介しよう。

2.2 例：MacLennan のサイモーグ

MacLennan (1990a,b) はコンピュータ内部の人工的環境にサイモーグ（simorg）という、単純な人工的生物を置いてシミュレーションを行い、それらがコミュニケーションを発達させる様子を観察した。この「世界」は、一つの大域環境と複数の局所環境からなっており、個々のサイモーグは共有の大域環境と固有の局所環境にアクセスを持つ。ある個体がアクセスを持つ局所環境には他の個体はアクセスを持たない。局所環境の状態は「シチュエーション」と呼ばれる。サイモーグはランダムに変化する局所環境の状態に反応して信号を発信する。大域環境の状態はサイモーグの発信する信号に応じて変化する。この大域環境の状態は「シンボル」と呼ばれる。シンボルを感知したサイモーグたちはそれに反応して何らかの行動を起こす。環境の状態、信号の種類、行動の種類はすべて 0 から 7 までの 8 個の整数によって表現される。

あるサイモーグがシチュエーションに応じた信号を発すると、その信号に応じてシンボルが変化する。シンボルに反応して他のサイモーグが起こした行動が最近の発信者のシチュエーションに一致するとき（これは発信者の信号に対して受信者が「協力的」な行動をとった、すなわち「コミュニケーション」が「成功」したものと考えられる）、発信者と受信者の双方に「得点」が与えられる。累積された得点がサイモーグの「適応度（fitness）」とみなされる。一定の周期で、サイモーグたちは適応度に応じた確率で、子孫を残し、あるいは死亡する（もちろん適応度の高い個体ほど子孫を残す確率が高く、適応度の低い個体ほど死亡する確率が高い）。各個体はまた、得点が得られた反応や信号を学習することも出来る。

実験は、信号が正常に伝達できる場合と伝達が抑制される場合、各個体が学習能力

測定値	伝達/学習		
	不可/不可	可/不可	可/可
α	6.31	11.63	59.65
$\dot{\alpha} \times 10^4$	0.36	11.0	28.77
η	0.94	0.26	0.29

表 1：MacLennan (1990b), Table 1 をもとに作成

を持つ場合とそうでない場合に分けて、合わせて 4 通りの場合で行われた。

新しい世代が誕生する周期を大周期として、一定数の大周期後の平均適応度 α と適応度の平均上昇率 $\dot{\alpha} \times 10^4$ が調べられた。またそれとは別に、一定数の大周期後に、コミュニケーションが行われた際の各個体が発信する信号とそれに対応するシチュエーションの組み合わせを行列で表したものが分析された。この行列は「表示行列（denotational matirix）」と呼ばれている。各エントリは、シチュエーションとシンボルの組み合わせに対応しており、その組み合わせによってコミュニケーションが行われる度にそのエントリの数字が 1 増やされる。

シンボルが全くでたらめに使われているとすれば、表示行列は各エントリの数が似通った数になるだろう。このような表示行列は「一様行列」と呼ばれる。逆にシンボルが有意味に使われているとすれば、表示行列の各行、各列には唯一つの 0 でないエントリが存在し、そしてそのエントリの数字はほぼ等しくなるだろう。このような表示行列は「理想行列」と呼ばれる。

記号がどれだけ有意味に使われているかどうかは、表示行列の分布が一様と理想の間のどこに位置しているかによって測ることが出来る。これは行列のばらつきの度合いとして、エントロピーを使って量化できる。この尺度を「無秩序測定値」と呼び、η で表す(5)。η は一様行列に対して 1、理想行列に対して 0、過度に構造化された対角行列に対しては –1 の値を持つ。5000 大周期後の測定値の平均は表 1 のようになった。

表示行列を分析すると、信号の伝達が阻害されている実験に比べて伝達が許可されている実験の方が行列が構造化されていた。多くの行、列において 2，3 個のエントリが突出して大きな数字になり、それ以外のエントリは 0 であるかもしくは比較的小さな数字であった。これは特定のシチュエーション－シンボルの組み合わせがコミュニケーションにおいて主に使われるようになっていることを意味する。また実験を通じて、ほぼすべてのシンボルが多義的に使われること、ほぼすべてのシチュエーションが複数の異なるシンボルによって意味されることが普遍的な現象として観察された。MacLennan によれば、このことは、多義語や同義語が自然言語にとって普遍的な特徴

であり、無視するべき現象ではない、ということを意味している(6)。

3 総合倫理学（SE）に向けて

総合的方法は言語やコミュニケーション、協力行動の発達などを研究する有力なツールであることが明らかになっている。本節ではこの方法がさらに道徳的行為の研究に活用することが可能か（そして有用か）どうかを検討しよう。

3.1 モデルと理論

前節で紹介した MacLennan のモデルは非常に単純なものであるが、それでも自然言語の発達に関して、いくつかの有意義な示唆を与えてくれる。このモデルの有意義性は、部分的には、記号の使用が意味のあるコミュニケーションになっていることが、表示行列の分析によって明らかになる、という前提に依拠している。より具体的にいえば、MacLennan は、サイモーグたちによる記号の使用がどれだけ意味のあるコミュニケーションになっているかは、表示行列の構造化の程度に反映される、と考えている。そしてその構造化の程度は、数値の分散の度合いとして数学的に定量的に測定することが出来る。

一般に、総合的動物行動学は抽象的なモデルによって現実の現象をシミュレートするため、そのモデルの特徴を現実の現象の特徴と関連付けるためには、単なる数値を適切に解釈して意味づける必要がある。シミュレーションによって現実の現象を研究する際には、モデルが現実の現象の良いアナロジーになっていなければならない。そしてモデルが良いアナロジーになっているかどうかを判定するためには、その基準となる理論が必要である。特に、モデルが現実の対象／現象のどの側面を反映し、どの側面を無視するべきかということについての説得力のある前提が必要なのである。

例えば風洞実験で重要なのは、モデル飛行機の翼の形状、重量、風速などであり、機体の色や素材は重要ではない。この背景には、揚力を決定するのは風速と翼の形状であり、機体の色や素材ではない、という前提がある。また経済学における一般的な市場モデルでは、商品の価格が生産・消費行動を決定する要因であり、品質の違いは無視される。この背景には、生産・消費行動の主要な要因は価格である、という前提がある。これらの前提にある程度の説得力があるからこそ、これらのモデルは現実を反映したモデルであるとみなされるのである。

総合的方法を倫理学に応用する際の困難の一つは、「倫理的である」という性質をモデルに帰属させるため基準をどう与えるか、という点にある。総合動物行動学の利点は、モデルの「透明性」にある。つまり研究されるモデルの様々なパラメータの値

が、観察者にとって常にアクセス可能になっている、ということである。しかしそのような可視的なパラメータによって、「意図」や「動機」などを表現することが出来るだろうか。もしもある行為の道徳性にとってこれらの概念が本質的であるならば、それをコンピュータやロボットでシミュレートすることは困難であるだろう。あるいは、少なくとも AI やロボティクスがそのようなシミュレーションを可能にする見込みは低い(7)。したがって、個人の意図や動機に道徳性の根拠を求める理論は、SE の基礎理論として適切ではない。

一方で、功利主義や帰結主義は、個人の意図や動機を問わず、それがどのような結果を及ぼすかということに基づいて行為の道徳性を判断する。しかし SE を基礎づけるためには、従来の帰結主義よりも、より客観的かつ定量的に道徳性を特徴づける必要がある。そこで、本稿では Floridi の情報倫理に注目し、それが SE のための有用な理論になりうるかを検討したい。

3.2 Floridi の「情報倫理」

近年、Floridi は「情報倫理」（以下 IE）という理論を提唱している。IE は以下の点によって特徴づけられる。

- 行為者よりも行為の結果の受容者（patients）を関心の中心とする。
- 人間や生物に限らず、すべての「情報的存在者」を道徳的受容者とみなす。
- 情報に関係する行為以上に、情報そのものに道徳的価値を置く。
- 情報的存在者の生息する領域、「情報圏（infosphere）」の情報の豊かさによって道徳性を規定する。

Floridi は、従来の倫理理論は情報社会の倫理、特に「コンピュータ倫理」（CE）のための一般的かつ首尾一貫した枠組みとして使うことが出来ない、と考える。その理由の一部として Floridi は、人間が自分の「行為者性、知性、自由、そして志向性（欲求、不安、期待、希望など）を計算システムに投射していること」、そして「ますます権威ある仲介者として、計算システムに責任を委任する傾向」を持つことを挙げている。行為者、そして行為を議論の中心に据える理論ではこのような問題に対処することが出来ない。にもかかわらず、CE の文献においては、従来の倫理学の不十分さを指摘する議論が見られない。これは CE が実践的な専門家の倫理としてしか見なされず、それゆえより哲学的に確立された倫理学に対して「非常に大きな劣等感」を持っていることの現れだ、と Floridi は言う(8)。そこで Floridi は、CE の基礎として適切な、理論的倫理学として IE を提唱するのである。Floridi によれば、

> ［...IE］はマクロ倫理学のスペクトラムが欠いている重要な役割を果たす。

> 私たちの倫理的な議論にはこれまで根本的な盲点が存在していた。IE と、応用の面で IE に対応する CE は、その盲点を知覚し、そして考慮に入れることが出来る。(Floridi, 1999, 55f)

IE の詳細に立ち入ること、IE を評価することは本稿の射程にはない(9)。ここでは簡単に、総合倫理学の基礎理論として IE を用いることの利点を検討したい。

第一に、IE はある行為に道徳性を帰属させる際に、行為者の意図や目的を問題にしない。問題になるのは、ある行為の結果として情報圏の状態がどのように変化したかである。これは総合的倫理学にとっては都合がよい。というのも、既に述べたように、もしも道徳性を判断するために行為者の意図を問題にしなければならないとすれば、人工的な行為者に意図や目的などを持たせることが可能かどうかが、まず大きな問題になるからである(10)。

第二に、より重要なことであるが、考慮されている情報圏の道徳的状態は、情報圏の持つ情報量に従って判定される。このとき、IE は「高度に発達した数学の分野（情報理論）に訴える」ことが出来る(11)。ただし Floridi は、そのような数学的手法によって現実の倫理的問題が解決できると考えるのは誤りだと言う。しかしコンピュータの中に作られた世界の道徳的状態を定量化して測定することが求められるのであれば、数学的なテクニックを使えるに越したことはない。というより、総合的手法にとっては、データの数学的分析が不可欠である。従って道徳性を情報量（エントロピーの低さ）によって規定する IE は、総合的方法にとって非常に便利な理論である。IE に対しては、現実の倫理問題に対処するには抽象的すぎる、あるいは形而上学的すぎるという批判がある(12)。しかしながら、抽象的な形而上学としての性格ゆえに、IE は倫理学への総合的アプローチを牽引する基礎理論としてのメリットを持つのである。

4 結論

本論は、Floridi によって提唱される情報倫理（IE）を、倫理学への総合的（synthetic）アプローチのために利用することの有用性を検討した。IE は行為者よりも受容者を道徳的関心の中心におき、また人間のみならずあらゆる情報的存在者を道徳的受容者として受け入れるという、普遍的な性格を持っている。また IE は情報的存在者の生息する世界（情報圏）の道徳的状態を、情報量という純粋に数学的な尺度によって測定することを提案する。これらの特徴は、総合的方法と相性が良い。もしも IE が倫理学の理論としてある程度の説得力を持つのであれば、IE に基づいて総合倫理学を構築する試みは意義があるだろう。

IE はまだ歴史の浅いものであり、その評価はこれから定まっていくものと思われる。しかし逆に総合倫理学への応用が IE の評価につながる可能性もある。いずれにせよモデルと理論は相補的なものである。理論はモデルにガイドラインを提供し、モデルは理論の有効性のテストを提供する。これまでの倫理学では、理論とモデルのこのような相互作用ということは考えられなかった。IE と総合動物行動学の結びつきは、倫理学に新しい研究方法を導入するかもしれない。

註

(1) Floridi (1999).
(2) 例えば Turing (1969), Searle (1980), Winograd & Flores (1986), Dennett (1992), Brooks (2006) などを参照。
(3) MacLennan (1990b) 参照。
(4) これは「遺伝的アルゴリズム」と呼ばれる。伊庭 (1994) 参照。
(5) η は

$$H = -\sum_k p_k \log p_k$$

$$\eta = H/\log N - 1$$

によって計算される。ただし N はシンボル（シチュエーション）の種類の数。H は分布のエントロピー。
(6) この実験結果のより詳細な分析については MacLennan (1990a,b) を参照。
(7) 久木田 (2008) 参照。
(8) Floridi (1999), 39.
(9) IE について、より詳しくは西垣通・竹之内禎 (2007) を参照。
(10) もっともこの問題自体は哲学的に興味深い問題である。例えば Dennett (1997) を参照。また久木田 (2008) も参照。
(11) Floridi (1999), 51 を参照。
(12) 例えば Rafael Capurro など。西垣通・竹之内禎 (2007) を参照。

文献

Brooks, R. (2006).『ブルックスの知能ロボット論 ── なぜ MIT のロボットは前進し続けるのか？』, オーム社. 五味隆志訳. *Flesh and Machines: How Robots Will Change Us*, 2002.

Dennett, D. C. (1992).「意識の進化とコンピュータの進化」,『〈意識〉の進化論』, 青土社, 33–63 頁. 斉藤健二訳. Evolution of Consciousness. In Brockman, editor, *Speculations: The Reality Club*, Prentice Hall Trade, 1990, pp. 85–108.

——— (1997). 'When HAL Kills, Who's to Blame?: Computer Ethics,' in Stork, D. G. ed. *HAL's Legacy: 2001's Computer As Dream and Reality*, Cambridge: The MIT Press, 351–365.

Floridi, L. (1999). 'Information ethics: On the philosophical foundation of computer ethics,' *Ethics and Information Technology*, *1*, 37–56.

MacLennan, B. (1990a). 'Evolution of communication in a population of simple machines,'Technical Report CS-90-99, Computer Science Department University of Tennessee, Knoxville.

——— (1990b). 'Synthetic Ethology: An Approach to the Study of Communication,'Technical Report CS-90-104, Computer Science Department University of Tennessee, Knoxville.

Searle, J. R. (1980). ‘Minds, brains and programs,’ *Behavioral and Brain Sciences, 1*, 417–424.
Turing, A. M. (1969). ‘Intelligent Machinery,’ in Meltzer, B. & Michie, D. eds. *Machine Intelligence, 5*: Edinburgh University Press, 3–23.
Winograd, T. & Flores, F. (1986).『コンピュータと認知を理解する ── 人工知能の限界と新しい設計理念』，産業図書．平賀譲訳．*Understanding Computers and Cognition*, 1986.
伊庭斉志 (1994).『遺伝的アルゴリズムの基礎 ── GA の謎を解く』，オーム社.
久木田水生 (2008).「ロボット倫理学の可能性」， *Prospectus*，第 11 巻，1–10 頁.
西垣通・竹之内禎 (2007).『情報倫理の思想』，NTT 出版.

〔京都大学非常勤講師／哲学〕